datacamp

COURSE GLOSSARY

Large Language Models (LLMs) Concepts

Artificial Intelligence (AI): The broad field of computer science focused on creating systems that perform tasks which normally require human intelligence, such as perception, reasoning, and decision-making

Attention Mechanism: A method that computes weights over input tokens to focus the model's computation on the most relevant parts of the input for a given prediction

Bag-of-Words: A simple text representation that encodes a document as counts or frequencies of unique words, ignoring word order and semantics

Bias (in data or models): Systematic skew or imbalance in training data or learned behavior that can produce unfair, inaccurate, or stereotyped outputs

Data Privacy / Personally Identifiable Information (PII): The principle and legal requirement to protect sensitive personal data used in model training and deployment to prevent misuse and comply with regulations

Deep Learning: A subset of machine learning that uses multi-layer neural networks to learn hierarchical and complex patterns from large datasets

Emergent Abilities: New capabilities that appear unexpectedly in large-scale models once they reach sufficient size in parameters and training data, enabling behaviors not present in smaller models

Hallucination: When a model produces confident but incorrect or fabricated information, often due to gaps or biases in its training data or reasoning process

Hallucination: When a model produces confident but incorrect or fabricated information, often due to gaps or biases in its training data or reasoning process

Lemmatization: The process of reducing words to their base or dictionary form (lemma) so that different inflected forms are treated as the same token

Machine Learning: A subfield of AI in which algorithms learn patterns and make predictions from data without being explicitly programmed for each task

Masked Language Modeling: A pre-training objective where certain tokens are masked and the model learns to predict those missing tokens from surrounding context to capture bidirectional representations

Model Explainability: Methods and properties that make a model's decisions and internal reasoning understandable to humans, improving trust and accountability

Multi-Head Attention: An attention design that runs multiple parallel attention operations (heads) to capture different types of relationships and combine them for richer representations

N-shot learning (zero/few/multi-shot): Transfer learning approaches where a pre-trained model is adapted to a new task using zero, a few, or many labeled examples respectively to enable performance with limited task-specific data

Natural Language Processing (NLP): The set of techniques and models that enable computers to understand, interpret, and generate human language in text or speech form

Next-word Prediction: A generative pre-training objective that trains a model to predict the next token in a sequence given preceding context, enabling coherent text generation

Parameters (model parameters): The internal numeric weights learned during training that encode patterns from data and determine a model's outputs

Positional Encoding: A technique to inject information about token order into transformer inputs so the model can distinguish different token positions in a sequence

Pre-training: The initial phase of training where a model learns general language structure and knowledge from massive, typically unlabeled text corpora using self-supervised objectives

Reinforcement Learning from Human Feedback (RLHF): An advanced fine-tuning technique where humans rank or rate model outputs and the model is optimized to produce outputs aligned with human preferences

Self-Attention: An attention variant where tokens in a single sequence attend to each other to capture contextual relationships among all tokens in that sequence

Stop Word Removal: A preprocessing step that removes very common words (e.g., "the", "is") that typically carry little semantic weight to reduce noise in text data

Text Representation: Methods to convert preprocessed text into numerical form that models can ingest, including sparse counts and dense vector embeddings

Tokenization: The process of splitting raw text into smaller units called tokens (words, subwords, or punctuation) to prepare text for computational processing

Transformer: A neural network architecture that processes input tokens in parallel using attention mechanisms and positional encodings to model long-range dependencies efficiently

Word Embeddings: Dense vector representations of words learned from data that place semantically similar words close together in vector space